

# Elimination of stripe artifacts in light sheet fluorescence microscopy using an attention-based residual neural network

ZECHEN WEI,<sup>1,2,3</sup> XIANGJUN WU,<sup>4</sup> WEI TONG,<sup>5</sup> SUHUI ZHANG,<sup>5</sup> XIN YANG,<sup>1,2,3</sup> JIE TIAN,<sup>1,2,6,7</sup>  AND HUI HUI<sup>1,2,3,8</sup> 

<sup>1</sup>CAS Key Laboratory of Molecular Imaging, The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Beijing 100190, China

<sup>2</sup>Beijing Key Laboratory of Molecular Imaging, Beijing 100190, China

<sup>3</sup>School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100190, China

<sup>4</sup>Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing 100083, China

<sup>5</sup>Senior Department of Cardiology, the Sixth Medical Center of PLA General Hospital, Beijing 100853, China

<sup>6</sup>Zhuhai Precision Medical Center, Zhuhai People's Hospital, affiliated with Jinan University, Zhuhai 519000, China

<sup>7</sup>tian@ieee.org

<sup>8</sup>hui.hui@ia.ac.cn

**Abstract:** Stripe artifacts can deteriorate the quality of light sheet fluorescence microscopy (LSFM) images. Owing to the inhomogeneous, high-absorption, or scattering objects located in the excitation light path, stripe artifacts are generated in LSFM images in various directions and types, such as horizontal, anisotropic, or multidirectional anisotropic. These artifacts severely degrade the quality of LSFM images. To address this issue, we proposed a new deep-learning-based approach for the elimination of stripe artifacts. This method utilizes an encoder-decoder structure of UNet integrated with residual blocks and attention modules between successive convolutional layers. Our attention module was implemented in the residual blocks to learn useful features and suppress the residual features. The proposed network was trained and validated by generating three different degradation datasets with different types of stripe artifacts in LSFM images. Our method can effectively remove different stripes in generated and actual LSFM images distorted by stripe artifacts. Besides, quantitative analysis and extensive comparison results demonstrated that our method performs the best compared with classical image-based processing algorithms and other powerful deep-learning-based destriping methods for all three generated datasets. Thus, our method has tremendous application prospects to LSFM, and its use can be easily extended to images reconstructed by other modalities affected by the presence of stripe artifacts.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

## 1. Introduction

Light sheet fluorescence microscopy (LSFM) is an optical sectioned fluorescence microscopic technique that can be used for fast and high-resolution imaging of biomedical samples with low photobleaching responses [1,2]. Currently, LSFM is extensively used in neurology [3], vascular analysis [4,5], and whole-body imaging [6]. However, stripe artifacts caused by high-absorption or scattering structures, such as impurities or bubbles, in the excitation light path or inside the sample result in image degradation in the application of LSFM [7].

Several techniques have been developed to eliminate stripe artifacts in LSFM images, which can be classified into hardware modification and image-based processing types [8]. In the cases of hardware modifications in LSFM systems, Dong et al. [9] improved the image quality

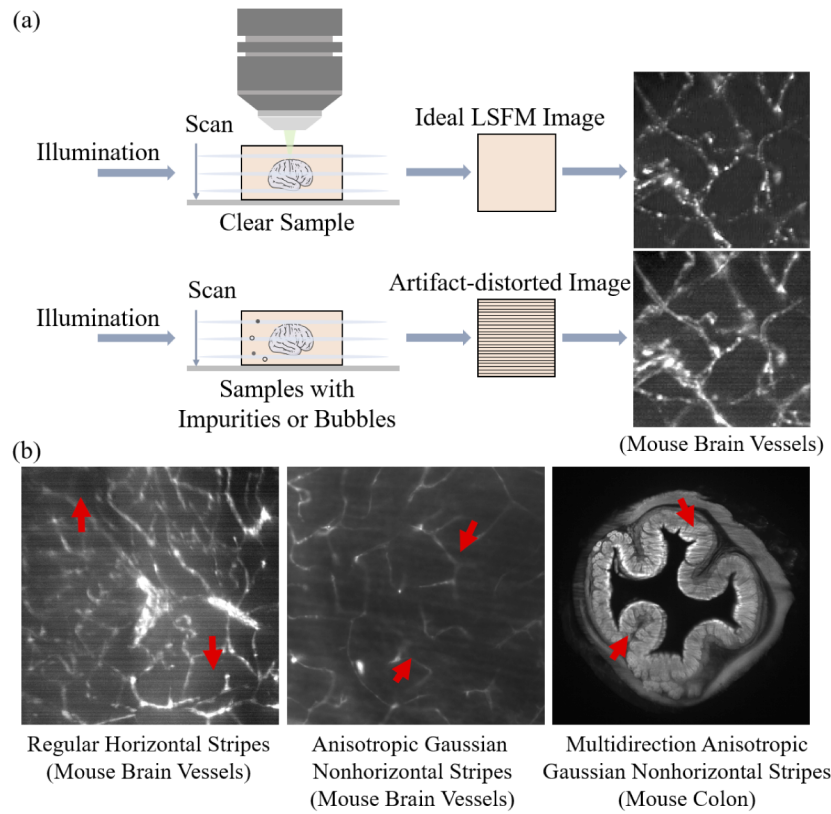
and reduced stripes in unidirectional LSFM by developing a vertically scanned LSFM method. Two parallel light sheets are used in the bidirectional LSFM system from both sides of the sample to eliminate the stripes [10]. Similarly, multidirectional LSFM averages images from different illumination directions and eliminates the stripes [11]. Ren et al. [12] proposed a novel approach called coded light-sheet array microscopy (CLAM), which allows complete parallelized 3D imaging without mechanical scanning, minimizing the illumination artifacts originated in highly scattering tissue. Besides, Ricci et al. [13] adopted acousto-optic deflectors (AODs) in LSFM to reduce the stripe artifacts in the images. Although modified LSFM systems can generate homogenous fluorescence excitation, additional hardware and alignment requirements will considerably increase the complexity of the LSFM system. More importantly, the modified LSFM system may still suffer from stripe artifacts in low-brightness or dense samples [6].

In addition to the aforementioned hardware modification techniques, several image-based processing algorithms have also been suggested to eliminate stripe artifacts [14,15]. Fehrenbach et al. [16] treated the elimination of stripes as a restoration problem and proposed the variational stationary noise remover (VSNR) method that removes stationary noise (such as that manifested in the form of stripes). In addition, Münch et al. [17] combined fast Fourier transforms and wavelet to filter out stripe noise. Liang et al. [18] designed a multidirectional stripe remover (MDSR) method, which applied a fast Fourier transform in the nonsubsampling contourlet transform domain to shrink the stripe components. Pollatou et al. [19] proposed a novel destriping method of stitched biological images based on the location of the stripe artifacts, background modeling, and illumination correction. Although these image-based algorithms can effectively remove stationary stripe artifacts, it is still challenging to remove artifacts that appear in random directions, as shown in Fig. 1.

Deep learning (DL) methods have been reported to significantly outperform conventional image processing algorithms in light microscopy, including microscopy image restoration [20], enhancement [21], super-resolution reconstruction [22,23], deconvolution [24], and microscopy image segmentation [25,26]. Wang et al. [20] proposed a UNet structure model to correct aberrations induced by the refractive index of fluorescence microscopy images. Instead of using a deeper convolutional neural network (CNN), Cai et al. [27] achieved state-of-the-art deblurring performance by utilizing deep residual networks (DRNs) to increase the network depth in dynamic motion deblurring for natural images. Abdallah et al. [28] proposed Res-CR-Net, which is a DRN, for the semantic segmentation of microscopic images. Although these pilot studies have shown promising results for image restoration and enhancement, deep-learning-based destriping of LSFM images has yet to be explored.

In this study, we propose an attention-based residual network (Att-ResNet) comprising residual blocks with a self-attention mechanism to efficiently eliminate stripe artifacts in LSFM images. We adopted the UNet architecture as the backbone of our network [29]. Residual blocks were implemented between successive convolutional layers of the UNet to address the blurring problem. To improve the performance of our network in the destriping task, a self-attention mechanism was introduced in the residual blocks. The self-attention mechanism comprises spatial and channel attention modules to learn sample features against stripe artifacts. Specifically, the spatial attention module aims to extract useful sample features from LSFM images. The use of channel attention is to exploit the features from different channels and remove the channels that originate from stripe artifacts according to the channel weights. Herein, spatial and channel attention modules were incorporated into the residual block to increase the efficacy of the proposed network.

To validate the performance of our model, we developed a degradation model to generate various stripe artifacts, including regular horizontal, anisotropic Gaussian nonhorizontal, and multidirectional anisotropic Gaussian nonhorizontal stripes for simulating stripe artifacts in



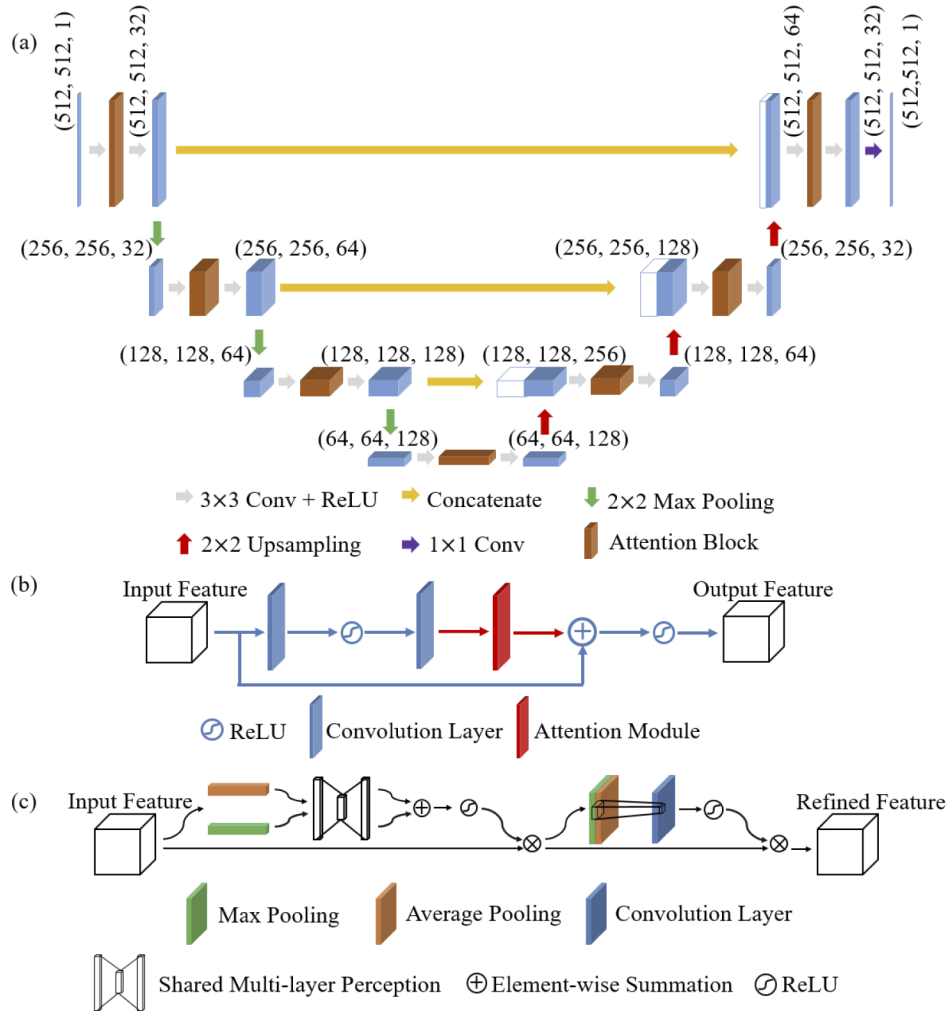
**Fig. 1.** (a) Illustration of the generation of stripe artifacts. (b) LSFM images are distorted by different stripe artifacts.

LSFM images. Both qualitative and quantitative comparisons with other state-of-the-art deep-learning-based and image-based algorithmic destriping techniques were conducted using the generated datasets. The results showed that the proposed method can effectively reduce various stripe artifacts. To the best of our knowledge, the proposed method is the first to address various stripe artifact issues in LSFM images with the use of a unified DL framework [8]. Moreover, our model was validated using external LSFM images of the mouse brain vessels [30], mouse colon [31], and mouse carotid plaques [32]. In addition, compared with the classic algorithm-based destriping methods, our method improves the processing speed by 60 times (compared with MDSR) and 180 times (compared with VSNR), which takes approximately 1s to process each image and thus enables fast destriping in LSFM images.

## 2. Methods

The main framework of our model was adopted from the UNet structure, as shown in Fig. 2(a). The encoder-decoder structure with skip connections allows the recovered feature map to integrate low-level features and aggregate features from different scales, so as to better distinguish the sample information and stripe artifacts in the image and obtain the ideal output. The skip connection integrates the features in the encoder and decoder through the concatenating of the channels of the feature. And the skip connection retains more dimensional information, so that the network can better learn from the abstract global features and the detailed local features. At

the same time, in the decoder part, upsampling will cause the loss of edge information of the feature map, and the edge feature can be retrieved by concatenating the feature from the encoder.



**Fig. 2.** Illustration of the structure of our method. (a) The main framework of the network consists of the encoding and decoding structures of UNet with an attention block in the middle of each layer. (b) Structure of the attention block is composed of the residual block (blue part) and attention module (red part). There are different types of attention modules, and their performances will be introduced and compared below. (c) Structure of the convolutional block attention module (CBAM).

The model includes four downsampling and upsampling operations. Each layer of the model consists of two  $3 \times 3$  convolution operations and an attention module in the middle. The feature dimensions of each layer were empirically determined by weighing the network performance and network complexity based on the existing related work [20,29,33], which were 32, 64, 128, and 256 respectively. Finally, a  $1 \times 1$  convolution layer transforms the resultant feature map from the last block to the output image.

### 2.1. Residual block

The motivation for adopting DRN is that the input (artifact-distorted LSFM images) and target output (ideal LSFM images) are expected to have similar values and structures. In addition, compared with plain CNN, residual networks contain identity mappings that prevent the gradient exploding or vanishing problem, thus facilitating the training of a deeper network. Therefore, to efficiently transform artifact-distorted images to artifact-reduced images, we used residual blocks to learn the difference between the input and target output. The residual block was composed of two convolution layers, a skip connection, and a rectified linear unit, as shown in Fig. 2(b).

### 2.2. Convolutional block attention module

To improve the model performance in the destriping task, we introduced an attention module that contained the self-attention modules, as shown in Fig. 2(c). This attention module can increase the modeling capacity of the network by assigning different weights to the features. Specifically, based on the importance of different feature maps, the value is selectively attenuated or amplified.

We set the attention module as  $\mathcal{F}$  and it works as follows,

$$\bar{I} = I \otimes \mathcal{F}(I), \quad (1)$$

where  $I \in R^{H \times W \times C}$  and  $\bar{I} \in R^{H \times W \times C}$  are the input and output of an attention module,  $H$ ,  $W$ , and  $C$  refer to the height, width, and channel depth of the feature map, respectively, and  $\otimes$  is an element-wise matrix multiplication. Existing attention modules can be divided into channel-wise and spatial-wise attention. Channel-wise attention obtains a vector  $v \in R^{1 \times 1 \times C}$ , which indicates the importance of each channel. Meanwhile, spatial attention calculates the importance of each pixel  $r \in R^{H \times W \times 1}$  to determine the most important regions.

Herein, we applied a convolutional block attention module (CBAM) [34] in our model (see Fig. 2(c)) and compared it with other attention modules, including global average pooling (GAP), squeeze-and-excitation module (SEM) [35], channel attention module (CAM), and spatial attention module (SAM). GAP is the simplest version of channel-wise attention and calculates the average of each channel of the feature map. SEM [35] is also a channel-wise approach that uses GAP as the squeeze operation, followed by two fully connected layers to model the interrelationship between channels. CAM and SAM are the channel submodules of the CBAM [34].

CBAM is composed of CAM and SAM as follows,

$$\mathcal{F}_{CBAM}(I) = I \otimes \mathcal{F}_{CAM}(I) \otimes \mathcal{F}_{SAM}(I \otimes \mathcal{F}_{CAM}(I)). \quad (2)$$

CAM utilizes GAP and global max pooling (GMP). GAP computes the channel-wise average value of the given input, while GMP returns the maximum value of the input feature map,

$$\mathcal{F}_{GMP}(I_c) = \max_{h,w} I_c(h, w), \quad 1 \leq h \leq H, 1 \leq w \leq W, \quad (3)$$

where  $I_c \in R^{H \times W}$ ,  $c \in \{1, \dots, C\}$  denotes the  $c^{th}$  feature of the input. The outputs of GAP and GMP are forwarded to a shared multi-layer perceptron, and the final output is calculated as follows,

$$\mathcal{F}_{CAM}(I) = \delta(MLP(\mathcal{F}_{GAP}(I)) \oplus MLP(\mathcal{F}_{GMP}(I))), \quad (4)$$

where  $\oplus$  refers to an element-wise summation, and  $\delta$  refers to sigmoid activation.

SAM encodes the importance of pixel-wise dependencies. Two pooling operations are employed to aggregate channel information,

$$s_{avg}(h, w) = \frac{1}{C} \sum_{c=1}^C I(h, w, c), \quad (5)$$

$$s_{max}(h, w) = \max_c I(h, w, c), \quad 1 \leq c \leq C, \quad (6)$$

where  $s_{max} \in R^{H \times W}$  and  $s_{avg} \in R^{H \times W}$  denote the spatial max pooling and spatial average pooling, respectively. The output of the pooling operations is then concatenated and passed through a convolutional layer and sigmoid activation to compute the SAM,

$$\mathcal{F}_{SAM}(I) = \delta(\text{Conv}([s_{avg} \oplus s_{max}]])), \quad (7)$$

where  $\oplus$  denotes the concatenating operation and Conv refers to a  $3 \times 3$  convolution layer.

### 2.3. Loss function

We aim to eliminate stripe artifacts while preserving useful sample information from LSFM images. To achieve this, we adopted the mean absolute error (MAE) and content loss [36] to design our loss function.

MAE calculates the average of differences between the pixels of the output and the ground truth

$$L_{MAE}(I, \hat{I}) = \frac{1}{H \times W \times C} \sum_{h,w,c} |\hat{I}_{h,w,c} - I_{h,w,c}|, \quad (8)$$

where  $I$  is the output of our model,  $\hat{I}$  is the ground truth label, and  $H$ ,  $W$ , and  $C$  refer to the height, width, and channel depth of the image, respectively. The difference at the pixel level is advantageous for the restoration of image intensity. However, MAE often ignores details, such as patterns, edges, and small variations.

Meanwhile, the content loss compares the feature map of the output and ground truth with the use of a pretrained feature extractor

$$L_{content}(I, \hat{I}) = \frac{1}{H \times W \times C} \sqrt{\sum_{h,w,c} (\phi_{h,w,c}(\hat{I}) - \phi_{h,w,c}(I))^2}, \quad (9)$$

where  $\phi$  denotes the feature map of the image. In our model, we used conv4-3 from the VGG-19 network [37] pretrained with ImageNet [38] as the feature extractor. This feature-level loss helped preserve content information and image details.

The final loss function is expressed as follows,

$$L_{total} = L_{MAE} + \lambda L_{content}, \quad (10)$$

where  $\lambda$  is a constant and set to be 0.05 in our model.

### 2.4. Evaluation metrics

The MAE, peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) were employed to evaluate the output image quality of the different methods. The MAE calculates the average per-pixel differences between the ground truth label and output. Thus, the smaller the MAE, the smaller the error between the ground truth label and the output. PSNR and SSIM are commonly used for image restoration and pay more attention to the fidelity of the image, which were



computed as:

$$MSE = \frac{1}{H \times W \times C} \sum_{h,w,c} |\hat{I}_{h,w,c} - I_{h,w,c}|^2, \quad (11)$$

$$PSNR = 20 \cdot \log_{10} \left( \frac{MAX}{\sqrt{MSE}} \right), \quad (12)$$

where  $MAX$  is the maximum intensity of image,  $I$  is the output of our model,  $\hat{I}$  is the ground truth label, and  $H$ ,  $W$ , and  $C$  refer to the height, width, and channel depth of the image, respectively.

$$SSIM = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)}, \quad (13)$$

where  $\mu_I$  and  $\mu_{\hat{I}}$  refer to the average value of  $I$  and  $\hat{I}$ ,  $\sigma_I$  and  $\sigma_{\hat{I}}$  refer to the standard deviation of  $I$  and  $\hat{I}$ ,  $C_1$  and  $C_2$  are two constants, and  $\sigma_{I\hat{I}}$  is the covariance of  $I$  and  $\hat{I}$ . The higher the PSNR, the better the image. The SSIM is a number that ranges from 0 to 1. The larger the result, the more similar it is to the original image.

## 2.5. Implementation

Our model was developed based on the TensorFlow and Keras packages. Att-ResNet was trained for 300 epochs using Adam optimizer with a batch size of two. The initial learning rate was set at 0.001. The angle parameters of the two classic algorithm-based methods for comparison were set according to a specific dataset, and the other parameters are set according to the optimal parameters in [16,18]. The training parameters of the other two DL-based methods were consistent with our network. In the training phase, only the DL methods were trained on the three training datasets ( $\sigma = 0.5$ ), while during the testing phase, DL-based and algorithmic methods were tested on the testing datasets constructed by different degradation models ( $\sigma = 0.3, 0.5, 0.7$ ). Our Att-ResNet can automatically remove artifacts, and it took approximately 1 s to process each image.

## 3. Experimental datasets

### 3.1. Data acquisition

All animal studies and procedures were performed according to a protocol approved by the Chinese People's Liberation Army General Hospital Animal Care and Use Committee in accordance with the National Institutes of Health Guideline on the Care and Use of Laboratory Animals.

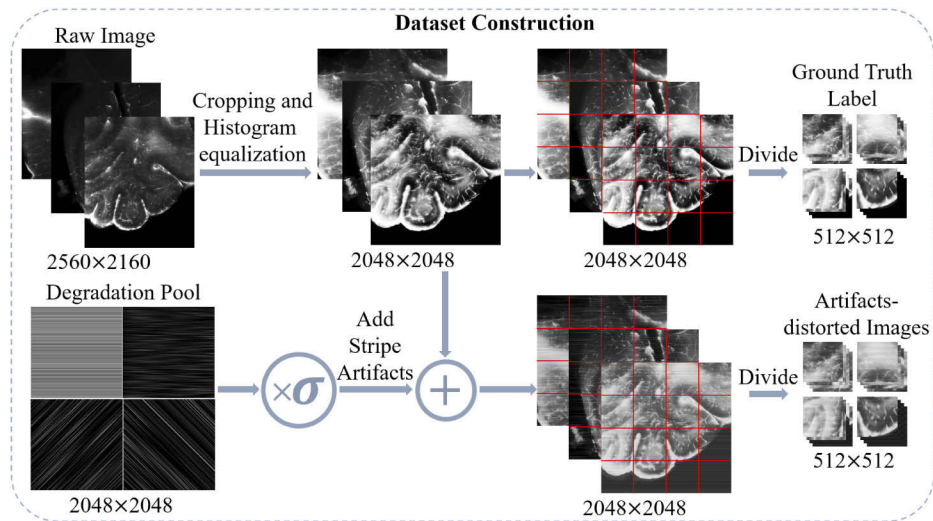
The process of immunolabeling and tissue clearing of mouse brains were the same as in [30]. The immunolabeled and cleared mouse brains were imaged on a commercial light sheet fluorescence microscope (Ultramicroscope II, LaVision Biotec, Bielefeld, Germany) with a 5.0× magnification, a 2× objective lens (Mv PLAPO2VC, Olympus), and a working dipping cap distance of 6 mm, which led to an image size of 2560 × 2160 with a pixel size of 0.65 μm. For the detection of brain vessels, the filters were set as follows: excitation 500/20 nm; emission 535/30 nm. The step size was set to 2 μm for z-stacks scanning and a total scan range of the brain sample up to 1 mm. The measurements were performed with exposure times of 385 ms per slice, resulting in a total acquisition time of ~2 min per brain resection sample. The raw LSFM images (saved in TIFF format) were processed using the ImageJ package FIJI (version 1.51, fiji.sc/Fiji, NIH, Bethesda, MD, USA). We collected coronal images of the resected mouse brain samples, including 2601 images of the cortex, 2061 images of the hippocampus, 2221 images of the cerebellum, and 1578 images of the olfactory bulb.

### 3.2. Degradation model

To verify the performance of our network, we developed a degradation model to generate various stripe artifacts to construct artifact-distorted image datasets [16,18]. We carefully designed the stripe artifacts of the datasets to simulate various types of stripe artifacts, including regular horizontal stripes, anisotropic Gaussian horizontal stripes, and anisotropic Gaussian nonhorizontal stripes. The first dataset was generated with regular horizontal stripes, anisotropic Gaussian horizontal stripes, and anisotropic Gaussian nonhorizontal stripes with angles of  $45^\circ$  or  $135^\circ$ . For the second and third dataset, we simulate the stripe artifacts in the horizontal direction, and the direction is randomly deflected based on the configuration of the commercial LSFM system. Thus, we generated the second dataset, including anisotropic Gaussian stripe artifacts with one random angle in  $[-11^\circ, 11^\circ]$ . The last dataset was generated using anisotropic Gaussian stripe artifacts with two angles randomly distributed in  $[-11^\circ, 11^\circ]$ . The bound of stripe angle in the second and the third datasets was chosen as per the configuration of the commercial LSFM system. All the images with stripe artifacts were randomly generated to form a degradation pool.

### 3.3. Dataset construction

The workflow of the dataset is shown in Fig. 3. To facilitate the subsequent operation, the raw image was first cropped to a size of  $2048 \times 2048$ , followed by image enhancement using a histogram equalization algorithm. Subsequently, artifact-distorted images were generated by randomly adding one of the artifacts from the degradation pool to the enhanced image.  $\sigma$  is a coefficient representing the intensity of the added artifact images. The artifact-distorted  $2048 \times 2048$  image was divided into 16 images with a size of  $512 \times 512$  pixels as the inputs. The ground truth labels were collected to initiate the training process.



**Fig. 3.** Workflow of dataset construction. We applied a series of preprocessing steps to the acquired LSFM image and added a degradation model to obtain the artifact-distorted images and the corresponding ground truth label, which were used to train our network.

For the training dataset, we extracted 15 slices cropped with  $2048 \times 2048$  in each of the collected images of the cortex, hippocampus, cerebellum, and olfactory bulb to determine the diversity of the dataset. After image division, we obtained 960 images of each degradation model ( $\sigma = 0.5$ ), including 864 and 96 images as the training and validation datasets, respectively. For the testing dataset, we extracted 60 slices from the collected images of the cortex, hippocampus,



and cerebellum, and three degraded image groups with different  $\sigma$  (0.3, 0.5, 0.7) comprised of 25920 subdivided images were generated as testing datasets.

## 4. Experiments and results

### 4.1. Ablation study

To study the effectiveness of our network, an ablation study was conducted to investigate the effects of the 1) position of the attention module, 2) different attention modules, and 3) content loss. All ablation experiments were performed on the second testing datasets ( $\sigma = 0.5$ ) with anisotropic Gaussian stripe artifacts with a random angle of  $[-11^\circ, 11^\circ]$ .

#### 4.1.1. Effect of the position of the attention module

We studied the model performance when the attention modules were added to different layers (@ ith), which refers to the attention modules located at the ith layer in a UNet architecture. ResNet refers to the standard UNet network with the addition of residual blocks between successive convolutional layers. The attention modules used in this experiment were CBAMs. The experimental results are summarized in Table 1.

**Table 1. Quantitative results of network outputs of different positions of the attention modules.**

	MAE ↓	PSNR ↑	SSIM ↑
Input	39.55±16.07	15.68±3.17	0.63±0.05
ResNet	6.35±2.15	30.02±2.46	0.89±0.08
@ 1st	5.94±4.33	31.12±2.98	0.92±0.07
@ 2nd	5.56±2.32	31.21±3.01	0.92±0.06
@ 3rd	5.98±2.37	30.49±2.67	0.89±0.07
@ 4th	5.65±2.30	31.14±2.81	0.90±0.09
@ All	<b>5.13±1.52</b>	<b>32.03±2.13</b>	<b>0.92±0.06</b>

In this experiment, we found that applying an attention module to any single layer could improve the model performance; this indicates the effectiveness of the attention module. In addition, we can see that applying an attention block to a layer had a similar effect on the model performance. By contrast, applying the attention module to every layer (@ All) achieved the best performance.

#### 4.1.2. Effects of different attention modules

We investigated the influences of the choices of different attention modules on a network based on ResNet. In this experiment, five networks were trained with different attention modules—GAP, SEM, CAM, SAM, and CBAM—as mentioned in Section 2.2. The experimental results are listed in Table 2. We observed that the networks with GAP, SEM, and CAM achieved significant gains in stripe artifact reduction compared with artifact-distorted LSF images. The performance of SAM is better than that of GAP, SEM, and CAM, indicating that the allocation of the spatial weight is more effective. The results of CBAM, which combined SAM and CAM, yielded the best performance compared with other attention modules. It is believed that CBAM can obtain the various important channels of feature maps and learn the information of different regions in the image to distinguish the sample information and stripe artifacts.

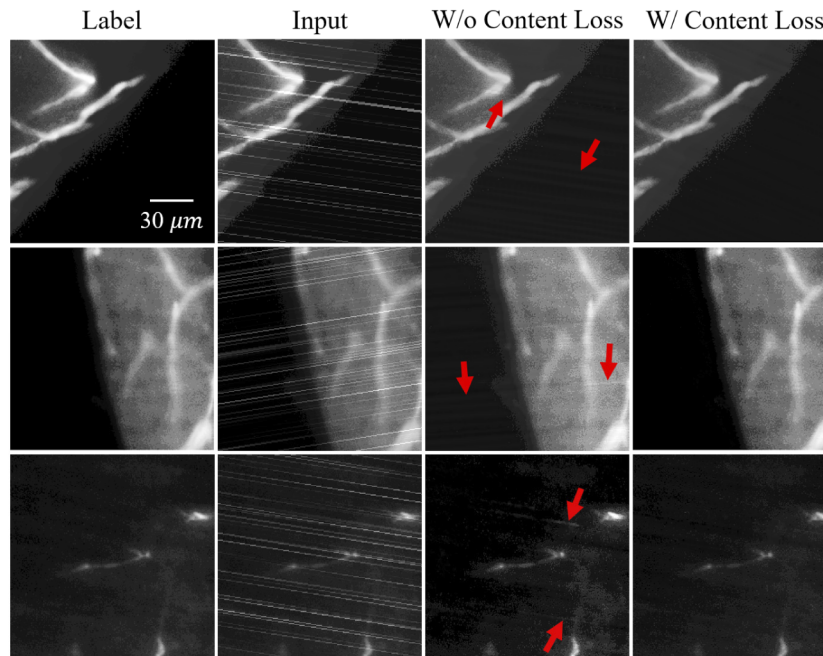
#### 4.1.3. Effects of content loss

In this comparative experiment, we studied the influences of the content loss on the model performance following the application of the MAE as the loss function. The comparative results

**Table 2. Quantitative results of network outputs of different attention modules.**

	MAE ↓	PSNR ↑	SSIM ↑
Input	39.55±16.07	15.68±3.17	0.63±0.05
GAP	5.89±2.18	30.95±2.75	0.91±0.05
SEM	6.59±2.29	29.96±2.61	0.91±0.05
CAM	6.64±1.98	29.60±2.21	0.89±0.05
SAM	5.63±2.30	31.13±3.09	0.91±0.06
<b>CBAM</b>	<b>5.13±1.52</b>	<b>32.03±2.13</b>	<b>0.92±0.06</b>

are presented in Table 3 and Fig. 4. The network with content loss achieved better performance than that without content loss in terms of MAE, PSNR, and SSIM. This is partly because the use of content loss can preserve the image details and content information. The benefit of the content loss can be clearly observed in Fig. 4. It is suggested that the use of content loss can reduce stripe artifacts in the input images while maintaining useful content information.

**Fig. 4.** Representative results obtained from the model with and without content loss. We selected three typical results as examples to elucidate the difference between results.**Table 3. Quantitative results of network outputs with and without content loss.**

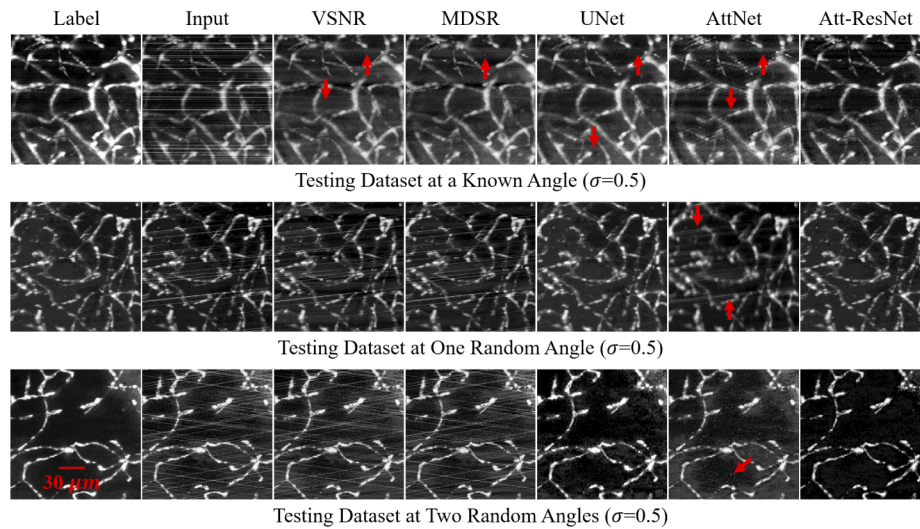
	MAE ↓	PSNR ↑	SSIM ↑
Input	39.55±16.07	15.68±3.17	0.63±0.05
w/o content loss	6.16±2.16	30.22±2.53	0.89±0.05
<b>w/ content loss</b>	<b>5.13±1.52</b>	<b>32.03±2.13</b>	<b>0.92±0.06</b>

#### 4.2. Comparison with existing destriping methods

We compared our proposed network with other existing classic algorithm-based and deep-learning-based destriping methods on different testing datasets of artifact-distorted LSM images. The following methods were chosen:

- VSNR: We chose the variational algorithm developed by Fehrenbach et al. [16], as it can be applied to the removal of stationary noise such as stripes from microscopy images.
- MDSR: We selected the algorithm reported by Liang et al. [18] because they designed a multidirectional stripe removal method following the application of the fast Fourier transform in the nonsubsampling contourlet transform domain to shrink the stripe components.
- UNet-based networks (UNet): The model proposed by Wang et al. [20] was chosen, as the UNet models are extensively used in designing networks for fluorescence microscopy image restoration.
- Self-attention mechanism-based networks (AttNet): The self-attention mechanism-based network (Ko et al. [39]) was selected because it achieved a noticeable performance in various artifact reduction tasks compared with other models.

The experimental results are summarized in Table 4 and Fig. 5. The performance of our network was better compared with those of the existing methods in both metrics and visual inspection.



**Fig. 5.** Representative results obtained from different destriping methods. We selected representative samples from the three testing datasets and showed the corresponding results obtained with different methods.

##### 4.2.1. Testing dataset at a known angle

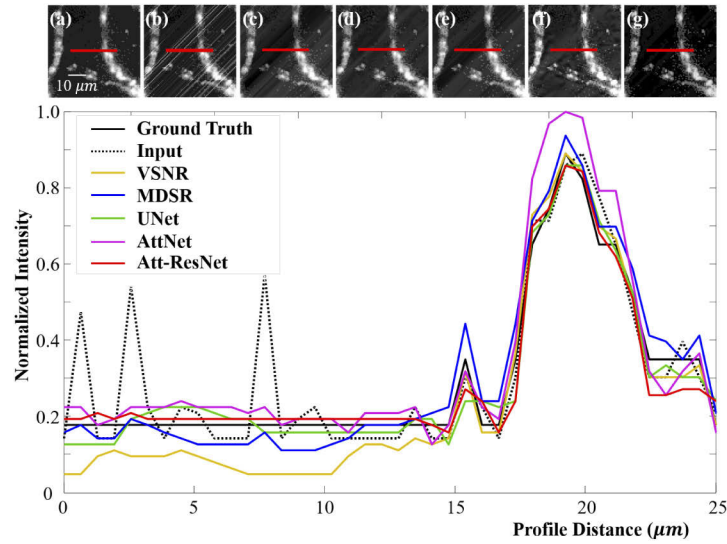
We first tested the effect of each method for handling stripe artifacts of a known angle. In this experiment, our Att-ResNet efficiently eliminates different stripe artifacts and achieves gains in the range of 0.33–4.23 in PSNR and 0.01–0.09 in SSIM ( $\sigma=0.5$ ). For the two classic algorithm-based methods, we set the angle parameters to 0, 45°, and 135° to deal with the different testing images, and they are effective for destriping tasks. PSNR of MDSR is 0.81 ( $\sigma=0.3$ ) and 0.29 ( $\sigma=0.7$ ) higher than our Att-ResNet. However, some visible artifacts are not

**Table 4. Quantitative results of outputs of our Att-ResNet and other destriping methods under different testing datasets.**

Method	Input	VSNR	MDSR	UNet	AttNet	Att-ResNet
<b>Testing Dataset at a Known Angle</b>						
$\sigma=0.3$ (n = 2880)	MAE ↓	25.40±10.50	6.51±2.78	5.57±2.54	6.03±1.78	10.77±4.87
	PSNR ↑	19.46±3.13	30.17±3.62	31.37±3.48	30.51±2.30	25.84±3.01
	SSIM ↑	0.64±0.11	0.90±0.10	0.92±0.08	0.82±0.08	0.83±0.07
$\sigma=0.5$ (n = 2880)	MAE ↓	36.10±14.85	7.59±3.13	6.37±2.69	6.84±3.35	10.25±4.34
	PSNR ↑	16.38±3.04	28.91±3.83	30.22±3.08	29.81±2.90	26.23±2.77
	SSIM ↑	0.49±0.11	0.86±0.12	0.88±0.10	0.89±0.08	0.81±0.06
$\sigma=0.7$ (n = 2880)	MAE ↓	44.39±18.32	8.73±3.55	6.35±2.27	6.75±1.67	10.55±3.95
	PSNR ↑	14.58±3.03	27.78±4.10	29.89±2.66	29.50±1.96	25.85±2.38
	SSIM ↑	0.39±0.10	0.81±0.14	0.85±0.11	0.83±0.08	0.78±0.06
<b>Testing Dataset at One Random Angle</b>						
$\sigma=0.3$ (n = 2880)	MAE ↓	29.02±11.02	7.74±0.08	7.97±2.14	5.77±3.37	6.04±2.26
	PSNR ↑	18.49±3.20	26.57±1.75	26.72±1.85	29.94±5.27	30.37±2.76
	SSIM ↑	0.74±0.09	0.75±0.08	0.76±0.07	0.89±0.12	0.90±0.08
$\sigma=0.5$ (n = 2880)	MAE ↓	39.55±16.07	12.37±2.54	10.47±2.21	6.81±2.70	6.60±5.28
	PSNR ↑	15.68±3.17	23.22±2.02	23.73±2.19	29.34±2.97	29.17±5.74
	SSIM ↑	0.63±0.05	0.65±0.08	0.65±0.08	0.87±0.08	0.85±0.12
$\sigma=0.7$ (n = 2880)	MAE ↓	50.48±19.31	13.40±1.98	12.90±2.60	8.22±3.09	7.38±6.92
	PSNR ↑	13.65±3.11	21.10±1.75	21.77±2.60	27.47±2.80	28.38±6.05
	SSIM ↑	0.53±0.07	0.54±0.06	0.58±0.08	0.80±0.08	0.81±0.11
<b>Testing Dataset at Two Random Angles</b>						
$\sigma=0.3$ (n = 2880)	MAE ↓	41.67±16.57	10.92±1.98	10.64±2.04	5.58±2.94	6.53±1.54
	PSNR ↑	15.36±3.34	24.66±1.56	24.51±1.60	30.52±4.43	29.76±2.28
	SSIM ↑	0.61±0.08	0.67±0.07	0.65±0.07	0.89±0.11	0.89±0.11
$\sigma=0.5$ (n = 2880)	MAE ↓	56.42±22.69	15.69±2.50	15.37±2.56	6.49±1.49	6.97±4.28
	PSNR ↑	12.73±3.35	21.38±1.51	21.14±1.56	29.73±2.19	29.33±2.16
	SSIM ↑	0.46±0.06	0.55±0.08	0.52±0.07	0.86±0.11	0.86±0.10
$\sigma=0.7$ (n = 2880)	MAE ↓	66.32±26.88	20.30±3.10	19.85±3.13	7.22±1.65	5.92±3.08
	PSNR ↑	11.34±3.37	19.16±1.49	18.90±1.60	28.64±1.95	29.82±4.45
	SSIM ↑	0.37±0.05	0.47±0.07	0.43±0.07	0.79±0.10	0.82±0.09

eliminated from the images, (see the red arrows in Fig. 5) and the consumption of each image requires approximately 180 s for VSNR and 60 s for MDSR. In addition, different parameters need to be adjusted when the two methods deal with different artifacts, thereby increasing the complexity of processing. By contrast, our method not only effectively removes different artifacts but also exhibits fast performance, as only 1 s is required to process one image. The performance of UNet and AttNet is comparable with that of our method in terms of PSNR and SSIM.

As shown in Fig. 6, the stripe artifacts cause the values of the original image to fluctuate violently, which seriously destroys the original structure. Our Att-ResNet can remove the artifacts and recover the distorted image to have a similar distribution as the ground truth label. MDSR and VSNR can restore images with remnant artifacts, especially those distributed in the background. The recovery ability of AttNet and UNet was slightly inferior to that of our method.



**Fig. 6.** Horizontal line profiles of the representative results ( $\sigma=0.5$ ) at row 378 and from column 380 to 420. The top row shows the enlarged region-of-interest of (a) the ground truth label, (b) artifact-distorted image (input), and results from (c) VSNR, (d) MDSR, (e) UNet, (f) AttNet, and (g) our method. The graph shows the normalized image values along the red lines on the top from left to right.

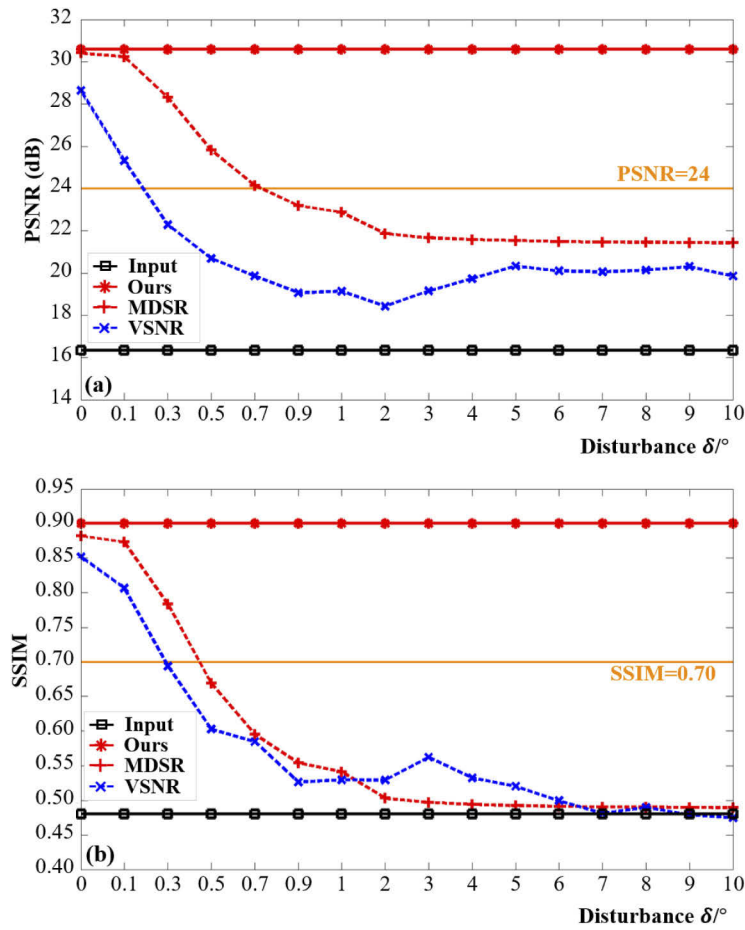
#### 4.2.2. Testing dataset at one random angle

In this experiment, we set the angle parameters to 0 for MDSR and VSNR to deal with the horizontal stripe artifacts with deflection that ranged from  $-11^\circ$  to  $11^\circ$  subject to the configuration of the commercial LSFM system. As shown in Fig. 5, the two classic algorithm-based methods cannot eliminate the stripe artifacts in the images because of the variation in the angle parameters that resulted in a significant decline in their performance. Our Att-ResNet achieves gains in the range of 5.07–10.02 in PSNR and 0.17–0.36 in SSIM compared with VSNR and MDSR methods. Our method also achieves the best performance, with gains in the ranges of 1.42–3.65 and 0.02–0.10 in terms of the PSNR and SSIM, respectively, compared with the UNet and AttNet methods. Besides, from Table 4, with the increase of  $\sigma$ , the performance of UNet and AttNet decreases (1.99–2.47 in PSNR and 0.09 in SSIM), while the performance of our Att-ResNet remains stable, which indicates the generalization performance of our method.

We studied the influence of angle parameters on the two classic algorithm-based methods, VSNR and MDSR, by adding disturbance  $\delta$  ( $\delta \in [0, 10^\circ]$ ) to the angle parameters and applying the two methods on the testing dataset at a known angle ( $\sigma=0.5$ ). The evaluation metrics of the results are shown in Fig. 7.

From Fig. 7, our Att-ResNet is not affected by the angle parameters (PSNR = 30.55 and SSIM = 0.90). When the angle parameters are accurate ( $\delta=0$ ), both VSNR and MDSR obtain satisfactory results (see Table 4). However, the performance of VSNR and MDSR decreases greatly when disturbance  $\delta$  is added to the angle parameter. When disturbance  $\delta$  is in the range of  $[0, 1^\circ]$ , their performance gradually decreases, and when it is in the range of  $[1^\circ, 10^\circ]$ , the two methods fail to remove the stripe artifacts. From Fig. 7, we can find that both the two classic algorithm-based methods are highly sensitive to the angle parameter, among which VSNR is more sensitive. When the disturbance  $\delta$  is more than  $1^\circ$ , these two methods cannot effectively remove the stripe artifacts in the image, indicating the high dependence of these two methods on the accuracy of the angle parameters.





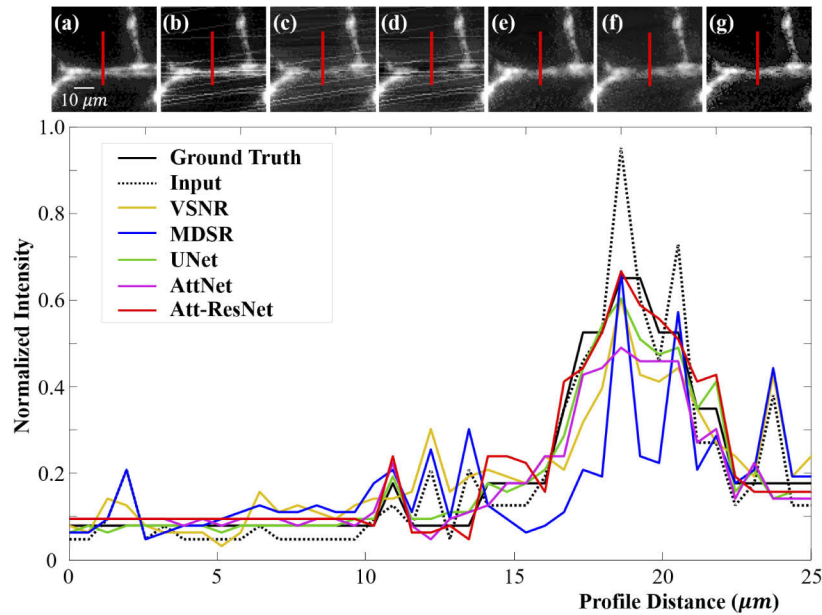
**Fig. 7.** PSNR and SSIM of the results under different disturbance  $\delta$  ( $\delta \in [0, 10^\circ]$ ). (a) is the PSNR of the testing dataset at a known angle, and (b) is the SSIM of the testing dataset at a known angle.

#### 4.2.3. Testing dataset at two random angles

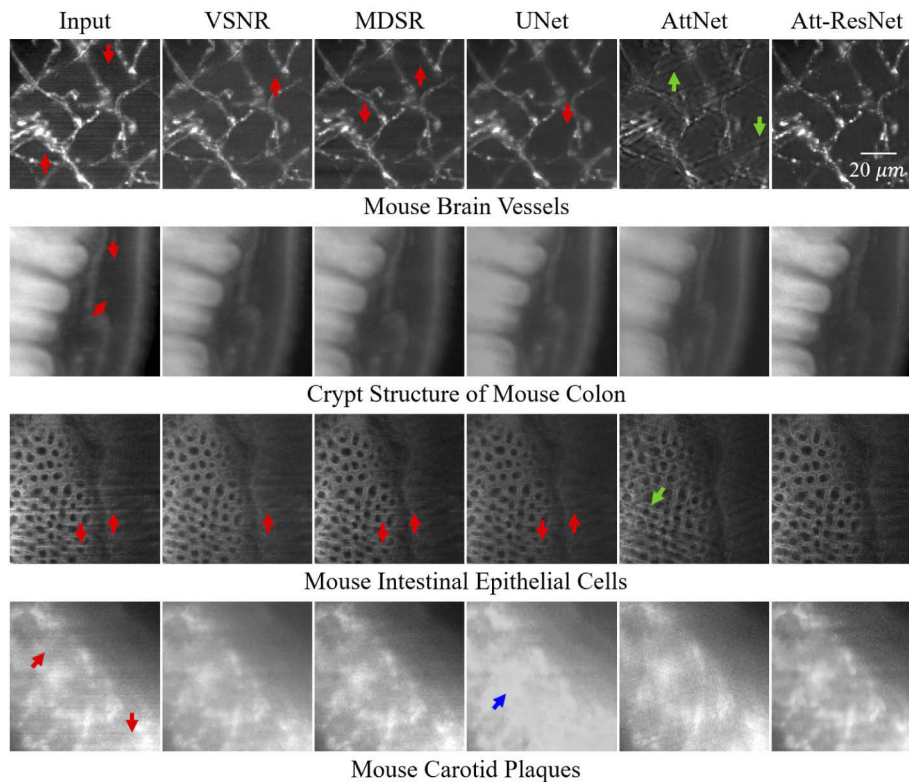
Finally, we applied the destriping methods to the testing dataset at two random angles, as shown in the last row of Fig. 5. The angle parameters of MDSR and VSNR were set as the previous testing dataset at one random angle. The two classic algorithm-based methods are still less effective in eliminating stripe artifacts. By contrast, the DL-based methods are successful in distinguishing stripe artifacts and samples and in eliminating artifacts at different angles in the images. However, some point noises appeared in the results of AttNet, which affected the overall quality of the image. In these testing experiments, stripe artifacts severely affected the image quality (PSNR and SSIM of the input were the lowest among the three testing datasets). Nonetheless, our model still effectively eliminated the artifacts, restored the image, and resulted in the best performance in both metrics and visual inspection.

We analyzed the line profile of each method on a testing dataset at two random angles ( $\sigma=0.5$ ). In Fig. 8, we found that the two classic algorithm-based methods are inefficient in the destriping task, and the DL-based methods yielded line profiles similar to the ground truth profiles compared with the classic algorithm-based methods. The line profile of our method is very close to the





**Fig. 8.** Horizontal line profiles of the representative results ( $\sigma=0.5$ ) at column 119 and from row 141 to 181. All remarks are the same as Fig. 6.



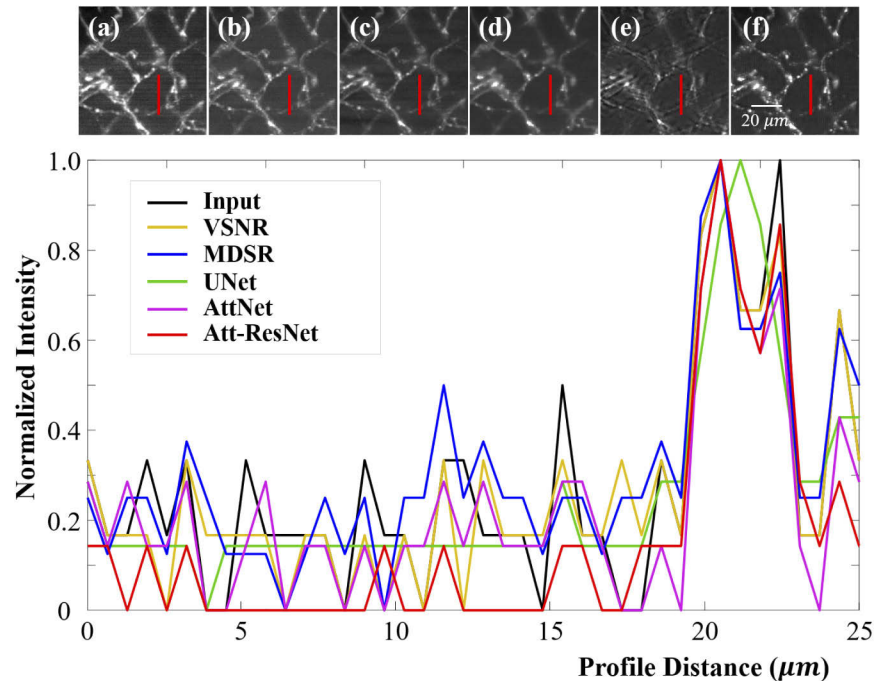
**Fig. 9.** Destriping results of LSFM images for different destriping methods.

ground truth, especially in the background, demonstrating that our method can restore the images in the presence of complex stripe artifacts.

#### 4.3. External testing on LSFM images with intrinsic stripe artifacts

In this section, we evaluate our Att-ResNet on LSFM images with intrinsic stripe artifacts in various biomedical applications. We validated our method using LSFM images with intrinsic stripe artifacts as the input, as shown in the first column of Fig. 9. As the stripe artifacts in the images are horizontal artifacts, we set the angle parameters of MDSR and VSNR to 0, and use the models trained by testing dataset at known angle for validation. From Fig. 9, all methods can effectively eliminate artifacts. For MDSR and VSNR, because of the accuracy of the angle parameters, the artifacts can be effectively removed. However, some visible artifacts are not eliminated from the images, (see the red arrows in the second and third column of Fig. 9). For UNet, in addition to some residual artifacts, it may blur the details of the images (see the blue arrow in Fig. 9). AttNet can effectively remove stripe artifacts, but it will introduce some additional ghosting artifacts into the image, which degrades the image quality (see the green arrow in Fig. 9). Our method can eliminate the stripe artifacts while preserving the image information.

For LSFM images with intrinsic stripe artifacts, due to the lack of ground truth for reference, we evaluated the performance of different methods by analyzing the line profile of each method of the representative LSFM image. In Fig. 10, the stripe artifacts cause the values of the original image to fluctuate violently, especially in the background (see the black curve). After destriping, the number of the spikes on the line profile will decrease, and the normalized image values of the background part will decrease and tend to 0. From Fig. 9(b)–9(f), all methods can



**Fig. 10.** Horizontal line profiles of the representative LSFM image with intrinsic stripe artifacts at column 1380 and from row 480 to 520. The top row shows the enlarged region-of-interest of (a) the original image, and results from (b) VSNR, (c) MDSR, (d) UNet, (e) AttNet, and (f) our method. The graph shows the normalized image values along the red lines on the top from up to down.

effectively remove artifacts, but some visible artifacts still remain in the result of VSNR and MDSR. Moreover, AttNet introduces some additional ghosting artifacts into the image. UNet and our Att-ResNet obtain a satisfactory result, reducing the number of spikes and the image values of the background part. Thus, UNet and Att-ResNet can better eliminate the artifacts while preserve the information of the image.

## 5. Discussion

The destriping of LSFM images is a challenging task that needs to be addressed by many recently developed methods. Both hardware improvements and classic algorithms have been proposed to eliminate stripe artifacts. However, hardware improvements still suffer from stripe artifacts in low-brightness or dense samples, which may affect the imaging speed and field of vision. For the two classic algorithm-based methods, VSNR and MDSR, the performance is satisfactory when dealing with artifacts with known angle (see Table 4). However, the two methods are sensitive to the angle of artifacts. And in practice, it is impractical to obtain all the angles of the artifacts from the LSFM images because the directions of the stripe artifacts are random and various, thus calculating all angles is time-consuming and complex. Moreover, the stripe artifacts are integrated into image. Therefore, artifacts cannot be effectively separated from the image information and measured accurately. When calculating the angle, the deviation of one pixel may cause an error of more than  $1^\circ$ . From Fig. 7, these two methods cannot effectively remove the stripe artifacts under this error. Although a more accurate result can be obtained by calculating multiple approximate angles, this can hardly be realized in practical application considering the time cost of the two methods ( $>60$  s per image), which require tremendous computational effort.

In this study, we proposed a DL-based method to effectively eliminate stripe artifacts from LSFM images. Unlike the classic destriping methods, our methods are based on the powerful prediction capability of deep convolutional networks to deal with the LSFM image distorted by different artifacts. In our Att-ResNet, we applied the residual blocks with the attention mechanism in our model to learn the features of the sample and different stripe artifacts in the image. We used residual blocks to increase the modeling power [40,41]. This learns only the difference between the input and the target output and transforms the artifact-distorted image to an artifact-reduced image efficiently. In addition, with CBAM, the importance of different regions and channels of the image was learned, and the value of the feature map was selectively attenuated or amplified according to their importance for more effective training. Moreover, CBAM increases the modeling capacity of the network to effectively eliminate artifacts and restore the image. We also use content loss to improve the performance of our model. Content loss is widely used in image processing algorithms of medical images, including artifacts elimination of CT images [39,42], super-resolution reconstruction of MRI images [43], and generation of synthetic digital mammography images [44]. In these works, a VGG network pre-trained on ImageNet is commonly used as the feature extractor to calculate the content loss. We carried out experiments on the content loss calculated by feature extractor pre-trained with ImageNet and LSFM images. The results can be found in Table S1 and Table S2 in Supplement 1. In Table S1 and Table S2, content loss can improve the performance of our model and the results of model pre-trained with ImageNet is better because the data scale and data diversity of ImageNet help the model to better extract the features in the image. Thus, in our model, we used conv4-3 from the VGG-19 network [37] pre-trained with ImageNet [38] as the feature extractor. This feature-level loss helped preserve content information and image details.

Our method was performed on the acquired images without the additional requirement of LSFM hardware improvement. In addition, Att-ResNet contributes to simple, fast image restoration and is suitable for images of different sizes. To verify the destriping ability of our Att-ResNet, we designed three degradation models to generate stripe artifacts in actual images to obtain the corresponding artifact-distorted images and ground truth LSFM image pairs. Our

Att-ResNet successfully restored stripe artifacts caused by the degradation models. Furthermore, our Att-ResNet was validated to eliminate the intrinsic stripe artifacts from the actual LSFM images. The performance of Att-ResNet was also quantitatively and qualitatively analyzed and verified to be better than the existing methods. The proposed method can be extended to eliminate different types of stripe artifacts in the LSFM images of ex vivo samples. Moreover, it is feasible to further improve the processing speed by optimizing our model, thus our method has the potential for processing in vivo imaging of neural activities and heart beating and can be applied to the processing of LSFM images in dynamic living samples. Moreover, it can be applied in the destriping task of other microscope images, such as scanning electron microscopy images, selective plane illumination microscope images, and FIB-nanotomography images.

Despite these advantages, our network remains highly dependent on data. The original image is regarded as the ground truth label in the constructed datasets, and artifacts and noise inevitably exist in the original image. The network can eliminate the artifacts and noise caused by the degenerate model. However, because of the presence of artifacts in the ground truth, some light artifacts in the original image cannot be eliminated completely. For the images with background noise, the noise will be regarded as effective information while removing artifacts, which would be remained as speckle noise. At the same time, some details in the image may be eliminated for being recognized as noise. This problem can be avoided by adding a degradation model to LSFM imaging of well-cleared samples to construct image pairs or by generating simulated LSFM images to avoid artifacts in ground truth labels. Moreover, a more complicated degradation model can be created to simulate complex noise, artifacts, and blur patterns in LSFM images. Additionally, a feature-extracted network can be designed pre-trained with available datasets to learn the sample information in the LSFM images of well-cleared samples and distinguish the noise information such as artifacts in the image. And the loss of the model can be upgraded using the designed network as a feature extractor to help improve the performance of main model, so as to eliminate the noise on the premise of preserving the image details as much as possible in practical application.

## 6. Conclusion

We proposed a DL method that combined a self-attention mechanism to eliminate stripe artifacts in LSFM images. Different degradation models were generated to simulate stripe artifacts in real LSFM images. Our method was validated by comparing it with two classic methods and two DL methods on the simulated data. The results showed that our method can effectively eliminate stripe artifacts in different datasets, thus overcoming the problem of angle sensitivity of classic methods for different stripe artifacts, and showing the ability to achieve fast performance. Moreover, our method was validated based on the destriping task on LSFM images with intrinsic stripe artifacts. Future studies will include the reduction of noise from different microscopic images and improvement in the temporal and spatial resolutions of the images.

**Funding.** National Key Research and Development Program of China (2017YFA0700401, 2016YFC0103803, 2017YFA0205200); National Natural Science Foundation of China (62027901, 81527805, 81671851, 81827808); Youth Innovation Promotion Association of the Chinese Academy of Sciences (2018167); Chinese Academy of Sciences Key Technology Talent Program; Project of High-Level Talents Team Introduction in Zhuhai City (HLHPTP201703).

**Acknowledgments.** The authors would like to acknowledge the instrumental and technical support of Multimodal Biomedical Imaging Experimental Platform, Institute of Automation, Chinese Academy of Sciences. We also thank Drs. Jianan Chen and Yongting Luo for providing the cleared brain samples.

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

**Supplemental document.** See [Supplement 1](#) for supporting content.



**References**

1. R. M. Power and J. Huisken, "A guide to light-sheet fluorescence microscopy for multiscale imaging," *Nat. Methods* **14**(4), 360–373 (2017).
2. T. Chakraborty, M. K. Driscoll, E. Jeffery, M. M. Murphy, P. Roudot, B. J. Chang, S. Vora, W. M. Wong, C. D. Nielson, H. Zhang, V. Zhemkov, C. Hiremath, E. D. De la Cruz, Y. T. Yi, I. Bezprozvanny, H. Zhao, R. Tomer, R. Heintzmann, J. P. Meeks, D. K. Marciano, S. J. Morrison, G. Danuser, K. M. Dean, and R. Fiolka, "Light-sheet microscopy of cleared tissues with isotropic, subcellular resolution," *Nat. Methods* **16**(11), 1109–1113 (2019).
3. E. A. Susaki, K. Tainaka, D. Perrin, F. Kishino, T. Tawara, T. M. Watanabe, C. Yokoyama, H. Onoe, M. Eguchi, S. Yamaguchi, T. Abe, H. Kiyonari, Y. Shimizu, A. Miyawaki, H. Yokota, and H. R. Ueda, "Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis," *Cell* **157**(3), 726–739 (2014).
4. R. Y. Cai, C. C. Pan, A. Ghasemigharagoz, M. I. Todorov, B. Forstera, S. Zhao, H. S. Bhatia, A. Parra-Damas, L. Mrowka, D. Theodorou, M. Rempfler, A. L. R. Xavier, B. T. Kress, C. Benakis, H. Steinke, S. Liebscher, I. Bechmann, A. Liesz, B. Menze, M. Kerschensteiner, M. Nedergaard, and A. Erturk, "Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull-meninges connections," *Nat. Neurosci.* **22**(2), 317–327 (2019).
5. W. Wang, Y. Q. Zhang, H. Hui, W. Tong, Z. C. Wei, Z. X. Li, S. H. Zhang, X. Yang, J. Tian, and Y. D. Chen, "The effect of endothelial progenitor cell transplantation on neointimal hyperplasia and reendothelialisation after balloon catheter injury in rat carotid arteries," *Stem Cell Res Ther* **12**(1), 9 (2021).
6. K. Tainaka, S. I. Kubota, T. Q. Suyama, E. A. Susaki, D. Perrin, M. Ukai-Tadenuma, H. Ukai, and H. R. Ueda, "Whole-body imaging with single-cell resolution by tissue decolorization," *Cell* **159**(4), 911–924 (2014).
7. A. Rohrbach, "Artifacts resulting from imaging in scattering media: a theoretical prediction," *Opt. Lett.* **34**(19), 3041–3043 (2009).
8. P. Ricci, V. Gavryusev, C. Müllenbroich, L. Turrini, G. de Vito, L. Silvestri, G. Sancataldo, and F. S. Pavone, "Removing striping artifacts in light-sheet fluorescence microscopy: A review," *Prog. Biophys. Mol. Biol.* **168**, 52–65 (2021).
9. D. Dong, A. Arranz, S. P. Zhu, Y. J. Yang, L. L. Shi, J. Wang, C. Shen, J. Tian, and J. Ripoll, "Vertically scanned laser sheet microscopy," *J. Biomed. Opt.* **19**(10), 1 (2014).
10. H. U. Dodt, U. Leischner, A. Schierloh, N. Jahrling, C. P. Mauch, K. Deininger, J. M. Deussing, M. Eder, W. Ziegler, and K. Becker, "Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain," *Nat. Methods* **4**(4), 331–336 (2007).
11. J. Huisken and D. Y. R. Stainier, "Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM)," *Opt. Lett.* **32**(17), 2608–2610 (2007).
12. Y. Ren, J. Wu, Q. Lai, H. M. Lai, and K. K. Tsia, "Parallelized volumetric fluorescence microscopy with a reconfigurable coded incoherent light-sheet array," *Light: Sci. Appl.* **9**(1), 1–11 (2020).
13. P. Ricci, G. Sancataldo, V. Gavryusev, A. Franceschini, and F. S. Pavone, "Fast multi-directional DSLM for confocal detection without striping artifacts," *Biomed. Opt. Express* **11**(6), 3111 (2020).
14. W. Ding, A. Li, J. Wu, Z. Yang, Y. Meng, S. Wang, and H. Gong, "Automatic macroscopic density artefact removal in a Nissl-stained microscopic atlas of whole mouse brain," *J. Microsc.* **251**(2), 168–177 (2013).
15. Y. Liu, J. D. Lauderdale, and P. Kner, "Stripe artifact reduction for digital scanned structured illumination light sheet microscopy," *Opt. Lett.* **44**(10), 2510–2513 (2019).
16. J. Fehrenbach, P. Weiss, and C. Lorenzo, "Variational algorithms to remove stationary noise: applications to microscopy imaging," *IEEE Trans. on Image Process.* **21**(10), 4420–4430 (2012).
17. B. Munch, P. Trtik, F. Marone, and M. Stampacchi, "Stripe and ring artifact removal with combined wavelet - Fourier filtering," *Opt. Express* **17**(10), 8567–8591 (2009).
18. X. Liang, Y. Zang, D. Dong, L. W. Zhang, M. J. Fang, X. Yang, A. Arranz, J. Ripoll, H. Hui, and J. Tian, "Stripe artifact elimination based on nonsubsampled contourlet transform for light sheet fluorescence microscopy," *J. Biomed. Opt.* **21**(10), 106005 (2016).
19. A. Pollatou, "An automated method for removal of striping artifacts in fluorescent whole-slide microscopy," *J. Neurosci. Meth* **341**, 108781 (2020).
20. L. Xiao, C. Y. Fang, L. X. Zhu, Y. R. Wang, T. T. Yu, Y. X. Zhao, D. Zhu, and P. Fei, "Deep learning-enabled efficient image restoration for 3D microscopy of turbid biological specimens," *Opt. Express* **28**(20), 30234–30247 (2020).
21. C. Bai, C. Liu, X. H. Yu, T. Peng, J. W. Min, S. H. Yan, D. Dan, and B. L. Yao, "Imaging enhancement of light-sheet fluorescence microscopy via deep learning," *IEEE Photon. Technol. Lett.* **31**(22), 1803–1806 (2019).
22. F. Zhao, L. X. Zhu, C. Y. Fang, T. T. Yu, D. Zhu, and P. Fei, "Deep-learning super-resolution light-sheet add-on microscopy (Deep-SLAM) for easy isotropic volumetric imaging of large biological specimens," *Biomed. Opt. Express* **11**(12), 7273–7285 (2020).
23. H. Zhang, C. Y. Fang, X. L. Xie, Y. C. Yang, W. Mei, D. Jin, and P. Fei, "High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network," *Biomed. Opt. Express* **10**(3), 1044–1063 (2019).
24. S. Qin, "Image reconstruction for large FOV Airy beam light-sheet microscopy by a 3D deconvolution approach," *Opt. Lett.* **45**(10), 2804–2807 (2020).
25. C. Kirst, S. Skriabine, A. Vieites-Prado, T. Topilko, P. Bertin, G. Gerschenfeld, F. Verny, P. Topilko, N. Michalski, M. Tessier-Lavigne, and N. Renier, "Mapping the fine-scale organization and plasticity of the brain vasculature," *Cell* **180**(4), 780–795.e25 (2020).

26. M. I. Todorov, J. C. Paetzold, O. Schoppe, G. Tetteh, S. Shit, V. Efremov, K. Todorov-Volgyi, M. During, M. Dichgans, M. Piraud, B. Menze, and A. Erturk, "Machine learning analysis of whole mouse brain vasculature," *Nat. Methods* **17**(4), 442–449 (2020).
27. J. R. Cai, W. M. Zuo, and L. Zhang, "Dark and Bright Channel Prior Embedded Network for Dynamic Scene Deblurring," *IEEE Trans. on Image Process.* **29**, 6885–6897 (2020).
28. H. Abdallah, A. Liyanaarachchi, M. Saigh, S. Silvers, and D. L. Gatti, "Res-CR-Net, a residual network with a novel architecture optimized for the semantic segmentation of microscopy images," *Mach. Learn.: Sci. Technol.* **1**(4), e8 (2020).
29. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.
30. J. N. Chen, Y. T. Luo, H. Hui, T. X. Cai, H. X. Huang, F. Q. Yang, J. Feng, J. J. Zhang, and X. Y. Yan, "CD146 coordinates brain endothelial cell-pericyte communication for blood-brain barrier development," *Proc. Natl. Acad. Sci. U. S. A.* **114**(36), E7622–E7631 (2017).
31. T. M. Li, H. Hui, C. E. Hu, H. Ma, X. Yang, and J. Tian, "Multiscale imaging of colitis in mice using confocal laser endomicroscopy, light-sheet fluorescence microscopy, and magnetic resonance imaging," *J. Biomed. Opt.* **24**(1), 1–8 (2019).
32. W. Tong, H. Hui, W. T. Shang, Y. Q. Zhang, F. Tian, Q. Ma, X. Yang, J. Tian, and Y. D. Chen, "Highly sensitive magnetic particle imaging of vulnerable atherosclerotic plaque with active myeloperoxidase-targeted nanoparticles," *Theranostics* **11**(2), 506–521 (2021).
33. W. B. Wang, B. W. Wu, B. Y. Zhang, X. J. Li, and J. B. Tan, "Correction of refractive index mismatch-induced aberrations under radially polarized illumination by deep learning," *Opt. Express* **28**(18), 26028–26040 (2020).
34. S. H. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of European Conference on Computer Vision* (Springer, 2018), pp. 3–19.
35. J. Hu, L. Shen, S. Albanie, G. Sun, and E. H. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(8), 2011–2023 (2020).
36. J. Johnson, A. Alahi, and F. F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proceedings of European Conference on Computer Vision* (Springer, 2016), pp. 694–711.
37. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
38. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "ImageNet: a large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition* (Springer, 2009), pp. 248–255.
39. Y. Ko, S. Moon, J. Baek, and H. Shim, "Rigid and non-rigid motion artifact reduction in X-ray CT using attention module," *Med. Image Anal.* **67**, 101883 (2021).
40. Q. Abbas, F. Ramzan, and M. U. Ghani, "Acral melanoma detection using dermoscopic images and convolutional neural networks," *Vis. Comput. Ind. Biomed. Art* **4**(1), 25 (2021).
41. L. Chen, R. Liu, D. Zhou, X. Yang, and Q. Zhang, "Fused behavior recognition model based on attention mechanism," *Vis. Comput. Ind. Biomed. Art* **3**(1), 1–10 (2020).
42. C. T. Peng, B. Li, P. X. Liang, J. Zheng, Y. Z. Zhang, B. S. Qiu, and D. Z. Chen, "A cross-domain metal trace restoring network for reducing X-Ray CT metal artifacts," *IEEE Trans. Med. Imaging* **39**(12), 3831–3842 (2020).
43. Q. Lyu, H. M. Shan, C. Steber, C. Helis, C. Whitlow, M. Chan, and G. Wang, "Multi-contrast super-resolution MRI through a progressive network," *IEEE Trans. Med. Imaging* **39**(9), 2738–2749 (2020).
44. G. F. Jiang, J. Wei, Y. S. Xu, Z. L. He, H. Zeng, J. F. Wu, G. G. Qin, W. G. Chen, and Y. Lu, "Synthesis of Mammogram From Digital Breast Tomosynthesis Using Deep Convolutional Neural Network With Gradient Guided cGANs," *IEEE Trans. Med. Imaging* **40**(8), 2080–2091 (2021).